Comparison of two methods for increasing training

Author/Contributor:
Milne, Linda; Willcock, Charles

Publication details:
Proceedings of the Eighth Australian Joint Conference on Artificial Intelligence, AI'95, Artificial Intelligence in the Environment Workshop
pp. 89-94

Event details:
AI in the Environment Workshop
Canberra, Australia

Publication Date:
1995

DOI:
https://doi.org/10.26190/unsworks/377

Comparison of Two Methods for Increasing the Training Set Size for Neural Networks

Linda Milne
linda@cse.unsw.edu.au

Charles Willock
charlesw@cse.unsw.edu.au

Computer Science and Engineering
University of New South Wales
Sydney NSW 2052
tel +61-2-385-3979
fax +61-2-313-7916

August 4, 1996

Abstract

The use of artificial neural networks with geographical data is often constrained by the very small number of training data points which can be obtained. Using multi-spectral and parametric data from the Nullica state forest in NSW, Australia, we look at addition of noise, and resampling, as methods of increasing the number and quality of the training set to get the most out of the data. Resampling the data appears to offer potential as a method of 'generalising' the neural network without the accuracy trade-off of added noise.

1 Introduction

Due to constraints of time, money and human resources, it is generally not possible to completely survey forests for species occurrence. However, it may be possible to survey a limited number of typical sites within a forest, and this data can be used to train classifiers such as artificial neural networks. The classifier can then be used to estimate species populations for the entire forest.

Neural networks are used extensively for pattern recognition problems, and are being adopted by the remote sensing and GIS communities (some examples include [1, 2, 3, 4]). Unfortunately neural networks require large amounts of training data to perform well. So, for applications such as vegetation mapping, where it is not always possible to obtain large amounts of training data we need ways of processing the data we do have so that the network can 'generalise' the known information to produce the best possible result. Neural networks also provide a powerful tool for analysing multi-source data [4], and so it is worth developing techniques for their use with small data sets.

One well known method for improving the generalisation capability of a neural network is to augment the training data set with clones of the training data points to which noise has been added [5]. Thus, we can increase the training set size by adding any number of noisy cases to the original data. However, if the data is already noisy it may in fact decrease the overall accuracy obtained.

Another possible method for improving the generalisation is to resample the data points. For each gridpoint in the data we can consider the value of each recorded parameter as defining a surface. Using cubic convolution interpolation [6] we can resample this surface to a smaller gridsize and, in the process, provide additional sample points – each slightly different from the original value. Resampling has the advantage of increasing the number of data points while potentially decreasing the noise in the dataset. It means that for each pixel location in the original training set we now have a cluster of pixels of similar value which have the same classification as the original value. This approach might well produce a better generalisation than the technique of adding random noise, as the feature values have been shifted in a way that retains the characteristics of the original environment.

In this paper we compare neural networks trained using data generated by these two methods with a neural network trained with the original data.

2 The Study Area

The area being studied is the Nullica State Forest on the south coast of New South Wales, Australia. The area is approximately $20km$ x $10km$ and is sampled on a grid of $30m$ x $30m$ pixels. The attributes used were aspect, slope, topographic position, altitude, geology, temperature and Landsat TM spectral bands 1 to 7. Surveys of the area have produced 190 training cases and 262 test cases. It is important to note that to generate considerably more training cases would be prohibitively expensive.

The data was normalised to fall in the range 0 to 1 to speed the rate of convergence of the network.

Initially we wish to obtain super type classifications for later use in more detailed species classifications. The classes being used are scrub (SC), dry sclerophyll forest (DS), wet/dry scleropyll forest (WD), wet sclerophyll forest (WS), and rainforest (RF).

A large proportion of the area is DS. The remaining classes have such small amounts of data that other methods will need to be used to obtain useful classifications. In this study we examine the classification accuracy for the classes DS and NOT DS.

3 The Neural Networks

A standard back-propagation network with 13 input units, 6 units in a single hidden layer and one output unit was used. The networks were trained to classify each input feature vector as DS or NOT DS. An input vector was given the target output value 0.9 if it was in the DS class and a target output value of 0.1 if it was not. For each of the training sets five copies of each network were trained with different initial weights. For each of the methods the performance figures obtained from the five networks were averaged to determine an overall performance figure.

The first method for increasing the training set size was to add noise to each of the input feature values [5]. For each of the normalised input vectors in the original training set three additional noisy vectors were added, giving a training set size of 760 vectors. This process of adding uniformly distributed noise was carried out to generate two training sets in addition to the original data. The first set had a noise range -0.01 to 0.01 (the 1% noise set), the second, a noise range -0.1 to 0.1 (the 10% noise set). The target output values remained unchanged.

The final training set was obtained using cubic convolution interpolation resampling [6]. Each attribute for the entire study area was resampled using a $15m$ x $15m$ grid. The four pixels corresponding to the original $30m$ x $30m$ pixel for the surveyed plot data were then extracted, again giving a training set size of 760 vectors. This data was also normalised to fall between 0 and 1.

The test cases are split into two sets. The first is used to prevent overfitting of the network and the second is used to obtain an unbiased assessment of network accuracy. Neither of the test sets were used for training the networks. When the total sum of squares error on the test set was at a minimum the training was stopped. The second test set was then used to measure the accuracy of the classifications.

The number of vectors in each class for each of the training and test sets is given in table 1. The noise and resampling techniques were not used for the test sets and, thus, all networks were assessed using the same test data.

	training set	test set 1	test set 2	random noise	resampled
DS	99	97	101	396	396
NOT DS	91	33	31	364	364

Table 1: Numbers of cases in each data set

4 Results

The number of correct classifications averaged over the five networks is given as a percentage in table 2 for each of the four paradigms.

	training set	test set 1	test set 2
original data	72%	59%	70%
original + 1% noise	76%	67%	73%
original + 10% noise	75%	67%	68%
resampled data	79%	67%	72%

Table 2: Average percentage of correct classifications

Table 3 shows the average percentage of cases from each class incorrectly classified by the trained networks.

	training set	test set 1	test set 2
original data	23%	34%	27%
original + 1% noise	8%	30%	22%
original + 10% noise	21%	33%	31%
resampled data	18%	29%	28%

(a) Average percentage of incorrect DS classifications

	training set	test set 1	test set 2
original data	35%	59%	42%
original + 1% noise	40%	61%	55%
original + 10% noise	29%	47%	45%
resampled data	23%	59%	38%

(b) Average percentage of incorrect NOT DS classifications

Table 3: Average percentage of incorrect classifications

5 Discussion

Examining table 2 we see an improvement in the percentage of correct classifications for the noisy and resampled training set data. This increase in accuracy was accompanied by an increase in the number of epochs to train the networks.

In table 2 we can see that the 1% noise dataset, and the resampled dataset provide a small increase in classification accuracy for test set 2. The corresponding 10% noise data set showed a marginal decrease in classification accuracy.

For the 1% noise dataset we see that the increase in the test set 2 accuracy is due to an decrease in the number of DS cases incorrectly classified (see table 3). This is in spite of the decrease in the accuracy for the NOT DS class.

For the 10% noise dataset, the incorrect classifications for both the DS and NOT DS classes (see table 3) appears to contribute to the slightly poorer classification performance seen in table 2. For this data there does not appear to be any consistent trend over the training and test sets. This could suggest that too much noise has been added.

In this study uniformly distributed random noise was added, for which the addition of noise at the 10% level seemed excessive. There could well be benefits in adding noise with a Gaussian distribution instead of the uniformly distributed noise. In that case the width of the distribution could reasonably be related to some parameter within the data.

For the networks trained on the resampled data the improvement comes from the ability to better recognise the NOT DS class, as seen by the decrease in the inaccuracy for the data in test set 2 (see table 3(b)). This result tentatively suggests that for small datasets resampling may offer benefits over the addition of noise as a way of 'generalising' the neural network.

6 Conclusions

The use of cubic convolution interpolation resampling allows us to increase the size of our training set, and slightly improve the 'generalisation' of the neural network when training with a size limited dataset. It does so without the potential penalty of a noise / performance trade-off seemingly inherent in the noise-added data expansion technique.

A detailed analysis of the use of resampling using neighbouring might yield additional benefits.

Improved pre-processing of the data to reduce quantisation effects and extract shadow effects in the data is also likely to improve results in future studies.

7 Acknowledgements

Linda Milne gratefully appreciates the financial support provided by Andrew Skidmore (School of Geography, University of New South Wales) and Brian Turner (Dept of Forestry, School of Resource and Environmental Management, Australian National University).

References

[1] H Bischof, W Schneider, and AJ Pinz. Multispectral classification of landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–

490, May 1992.

[2] T Yoshida and S Omatu. Neural network approach to land cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1103–1109, September 1994.

[3] AK Skidmore, W Brinkhof, and J Delaney. Using neural networks to analyse spatial data. In *Proceedings of the 7th Australasian Remote Sensing Conference*, pages 235–246, 1994.

[4] DZ Sui. Recent applications of neural networks for spatial data handling. *Canadian Journal of Remote Sensing*, 20(4):368–380, December 1994.

[5] J Sietsma and RJF Dow. Creating artificial neural networks that generalise. *Neural Networks*, 4:67–79, 1991.

[6] JA Richards. *Remote Sensing Digital Image Analysis : An Introduction.* Springer-Verlag, 2nd ed. edition, 1993.